

The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis

FRIEDRICH LÖSEL and MARTIN SCHMUCKER*

University of Erlangen, Nuremberg, Germany

*corresponding author: E-mail: mnschmuc@phil.unirrlangen.de

Abstract. The article reports a meta-analysis on controlled outcome evaluations of sexual offender treatment. From 2,039 documents published in five languages, 69 studies containing 80 independent comparisons between treated and untreated offenders fulfilled stepwise eligibility criteria (total $N = 22,181$). Despite a wide range of positive and negative effect sizes, the majority confirmed the benefits of treatment. Treated offenders showed 6 percentage points or 37% less sexual recidivism than controls. Effects for violent and general recidivism were in a similar range. Organic treatments (surgical castration and hormonal medication) showed larger effects than psychosocial interventions. However, this difference was partially confounded with methodological and offender variables. Among psychological programs, cognitive-behavioral approaches revealed the most robust effect. Nonbehavioral treatments did not demonstrate a significant impact. There was no outcome difference between randomized and other designs, however, group equivalence was associated with slightly larger effects. Various other moderators had a stronger impact on effect size (e.g., small sample size, quality of outcome reporting, program completion vs. dropout, age homogeneity, outpatient treatment, and authors' affiliation with the program). More differentiated, high-quality evaluations are needed to clarify: What works for whom under which circumstances?

Key words: evaluation, meta-analysis, recidivism, sex offender treatment, treatment efficacy

Abbreviations: CG = comparison group; TG = treatment group; ES = effect size; OR = odds ratio; LOR = logged odds ratio; $CI_{95\%}$ = 95% confidence interval

Treatment for sexual offenders is a very important topic of criminal policy. Media reports on serious cases of sexually motivated murder, rape, and child abuse have made people particularly concerned about this area of crime. In various countries, policymakers have reacted by increasing measures of both punishment and treatment. Because most incarcerated sexual offenders return to the community, effective treatment is a cornerstone for preventing future offenses. However, the empirical basis of sex offender treatment is less solid than such a cornerstone should be. Although recent overviews suggest a moderately positive effect (e.g., Alexander 1999; Aos et al. 2001; Gallagher et al. 2000; Hanson et al. 2002; Lösel 2000; Polizzi et al. 1999), methodological problems, inconsistent results, and a lack of high quality studies question how far we know what works for sex offenders (e.g., Lösel 2000; Marques et al. 2005; Quinsey et al. 1993; Rice and Harris 2003; White et al. 1998).

Indeed, there is much less well-controlled research on the evaluation of programs for sex offenders than in the field of general offender treatment (Lipsey and Wilson 1998; Lösel 2001a; McGuire 2002). Clear messages on the efficacy of programs for

sex offenders are complicated by various problems that relate to issues of design and treatment.

Design issues. Serious offenders relatively often receive at least some kind of treatment. This makes it difficult to form randomized untreated control groups. The relatively low base rate of official sexual reoffending (e.g., Hanson and Bussière 1998) means that large samples are needed to reveal a significant effect. However, for economic reasons, evaluations of sex offender treatment often contain rather small samples. This enhances the risk of nonequivalent groups, even in randomized designs (Marques et al. 2005). Furthermore, selective dropout and other problems make it difficult not only to implement but also to maintain internal validity (e.g., Barbaree 1997; Marshall and McGuire 2003). In addition to the general problems of undetected crime, the registration of sexual offenses is influenced by the intimate nature of this kind of crime. This limits the value of official data.

Treatment issues. The analysis of sexual reoffending requires relatively long follow-up periods (Prentky et al. 1997), so that evaluations may address old programs that do not represent the current 'state of the art.' Furthermore, the treatment of sex offenders embraces a broad range of medical and psychosocial programs, and sex offenders are a rather heterogeneous group. As a result, adequate evaluations and replications need to be rather differentiated. Moreover, sex offenders often do not receive a single program but individualized 'packages' (e.g., psychological and pharmacological treatment; additional programs for alcohol dependence). Disentangling the impact of specific modules would require relatively complex evaluations.

These and other problems make systematic research integrations more difficult than in those areas in which we can draw on a large number of methodologically sound evaluations. Accordingly, an early meta-analysis from Furby et al. (1989) found few well-designed studies and concluded that there is "as yet no evidence that clinical treatment reduces rates of sex reoffenses" (p. 27). Nine years later, a review for the Cochrane Collaboration came to a similar conclusion (White et al. 1998). Restricting their analysis to randomized evaluations, the authors ended up with only three studies. Other recent meta-analyses included quasi-experimental studies as well. This is in line with the notion that methodological quality is not a unidimensional all-or-none category, and that practice can also learn from good but suboptimal studies (Cronbach et al. 1980; Lösel and Köferl 1989; Shadish et al. 2002).

Most reviews report a lower average rate of sexual recidivism in treated groups than in control groups. For example, Hall (1995) integrated 12 controlled studies. The average rate of sexual recidivism was 19% in treated groups and 27% in controls (mean effect size (ES): $d = 0.24$). Hormonal and cognitive-behavioral treatments were more successful than other programs. However, such differential effects based on a small number of studies and a single evaluation of castration had a strong impact. Alexander (1999) integrated 79 studies on psychosocial sex offender treatment. The mean difference in recidivism was 5 percentage points in favor of treatment ($d = 0.12$; Lösel and Schmucker 2003). However, the majority

of studies contained no control group (Level 1 on the Maryland Scale of Methodological Rigor; Sherman et al. 1997). The same holds for a meta-analysis of 20 studies on medical and psychological treatment from German-speaking countries. This revealed a nonsignificant effect of $d = 0.08$ (Lösel 2000). Gallagher et al. (2000) analyzed 23 studies of relatively good quality. Treated groups showed 10 percentage points less sexual recidivism than controls and the overall effect size was relatively large ($d = 0.47$). In the most comprehensive recent meta-analysis, Hanson et al. (2002) integrated 43 studies on psychological treatment. The average sexual offense recidivism was 12.3% for treatment groups and 16.8% for comparison groups (mean ES: $d = 0.13$; Lösel and Schmucker 2003). Current, mainly cognitive-behavioral treatment revealed better outcomes whereas older programs appeared to have little effect.

Overall, the last decade has shown a strong increase and more positive outcomes in evaluations of sex offender treatment. However, research syntheses vary in effect size, type of treatment included, prevailing design quality, categorization of programs, treatment settings, and meta-analytical techniques (Lösel and Schmucker 2003). Nearly all are restricted to English-language studies that mainly come from North America. Against this background, an updated, comprehensive, independent, and international review seems worthwhile. This is why the Campbell Collaboration Group on Crime and Justice (Farrington and Petrosino 2001) has included a systematic review on sex offender treatment in its portfolio. The present study is a first report of our work on this synthesis. It contains a meta-analysis of studies on both psychological and biological treatment that were published up to 2003 in English and various other languages.

Method

Criteria for inclusion of studies

Primary studies had to have the following characteristics to be eligible:

1. *Study of sexual offenders.* Participants had to have been convicted of a sexual offense or to have committed acts of illegal sexual behavior that would have lead to a conviction if officially prosecuted.
2. *Evaluation of treatment.* No restrictions were made on the kind of intervention applied as long as it aimed to reduce recidivism (i.e., psychosocial as well as organic treatment modes such as hormonal medication or surgical castration were included). However, interventions had to incorporate therapeutic measures; purely deterrent or punishing approaches were not included. Treatment did not have to be specifically tailored for sexual offenders. More general measures of offender treatment were included if the study addressed at least a subgroup of sexual offenders and reported separate results for sexual offenders in both the treated and control groups.

3. *Measure of recidivism as outcome.* Recidivist behavior had to be included as a dependent variable. We followed a broad definition of recidivism and included outcomes ranging from incarceration to lapse behavior. In contrast, primary studies focusing exclusively on changes in measures of personality or hormone levels, clinical ratings of improvement, and the like were not included.
4. *Control group design.* The study had to report the same recidivism outcome for a comparison group not receiving the same treatment. This could either be an untreated control group or a group of offenders receiving treatment 'as usual' or another kind of treatment that differed from the evaluated program in content, intensity and specificity. Studies using only treatment dropouts as a control group were not eligible.
5. *Sample size.* Studies had to contain a minimum total sample size of 10 persons with at least five offenders in each group. In very small samples, even minimal differences in absolute numbers of recidivism between groups may lead to extreme effect sizes, thus making results very heterogeneous while not being very reliable.
6. *Sufficient data for effect size computation.* Studies had to report outcomes in a way permitting the calculation of effect size estimates.
7. *Country of origin.* No restrictions were made as to where studies were conducted. For economic reasons, we restricted our analysis to studies reported in English, German, French, Dutch, or Swedish.
8. *Time of publication.* There were no restrictions regarding the time of publication. All studies reported up to June 2003 were eligible.

Literature search

Searches were designed to tap published as well as unpublished evaluations. A wide range of sources was used to identify relevant studies. First, a basic study pool was compiled from the reference sections of previous reviews. Second, already identified primary studies were analyzed for further references. Third, 14 electronic databases were analyzed (e.g., PsycInfo, MedLine, ERIC, Cochrane Library, Social Services Abstracts, NCJRS abstract and full-text databases, Dissertation Abstracts International, and UK National Health Service National Register). Fourth, we hand searched journals pertaining to the topic (e.g., Sexual Abuse, Journal of Sexual Aggression). To identify more unpublished work, we contacted researchers in the field of sexual offender treatment and asked if they knew or had personally conducted further evaluations. Finally, an Internet search was conducted. Because the Internet constitutes a vast pool of rather loosely connected information, it cannot be searched in total. We visited the Internet sites of pertinent institutions and Departments of Corrections and searched them and their links for relevant material.

A total of 2,039 citations were identified in this way. These were scanned by their title first. If not obviously off-theme, we then examined the abstracts. If these

indicated potential eligibility or if no abstract was available, we used the full reports to finally evaluate eligibility. Twenty-one (1%) of the reports could not be obtained (of which 18 also had no available abstract). A total of 586 citations were clearly off-topic; 242 referred to offenders in general; and 641 did not evaluate a program, that is, they either simply described a treatment or reviewed sex offender treatment more generally. This left 549 citations referring to primary studies evaluating sexual offender treatment. Of these, 189 were excluded because they did not report recidivism outcome, 236 did not employ an adequate comparison group, 56 duplicated evaluations already included, and two did not lend themselves to effect size calculation. In the end, 66 reports met the specified inclusion criteria. Some contained more than one eligible study. In such cases, we referred to the individual studies as the unit of analysis. Some primary studies presented results for different subgroups (e.g., offense types). To allow for maximum differentiation while adhering to the principle of independency between effect sizes, we chose these subgroups as units of analysis. However, the characteristics of the subgroups had to be reported in just as much detail as the total sample. Following these rules, we formed a database of 80 comparisons from 69 studies.

Coding of study characteristics

Coding followed a detailed manual. A selection of variables is presented in the Results section (see Table 1). The manual covered general characteristics of the publication (e.g., year, country, type of publication), sample description (e.g., sample size, types of offender, age, voluntariness of treatment participation), methodological features (e.g., study design, type of control group, follow-up interval, source of recidivism data), and characteristics of the treatment (e.g., mode, setting, integrity, and format of treatment). Various programs for sex offenders are not restricted to a certain therapeutic paradigm but combine strategies from different 'schools' in an eclectic manner. Thus, as well as categorizing the basic therapeutic approach, we also rated the degree to which the different treatment elements were applied in an individual intervention on separate 4-point scales.

We evaluated the overall methodological quality of the individual study with an integrative rating scheme (see, for various options, Farrington 2003). We adapted the Maryland Scale of Scientific Rigor (Sherman et al. 1997) for our purposes. This is a 5-point scale integrating methodological features related to the validity of a causal interpretation of treatment effects. The highest level (5) is reserved for uncompromised random designs. Level 4 covers studies applying procedures to ensure group equivalence (e.g., individual matching, statistical control) or slightly compromised random designs. Designs based on incidental assignment are on Level 3 if group equivalence can be assumed (e.g., demonstrated equivalence on relevant variables). Studies incorporating a nonequivalent control group correspond to Level 2. Finally, Level 1 is reserved for uncontrolled studies that were not eligible for our meta-analysis. In its original form, the Maryland Scale also covers sample size and adequacy of statistical testing. However, because we integrated

studies by means of weighted effect sizes, these aspects were not so relevant and we dropped them in our rating.

Following Lösel and Köferl (1989), we included ratings on the completeness and accuracy of information reported. On a 4-point scale, these measured threats to 'descriptive validity' in the domains of treatment concept, treatment implementation, assessment of treatment goals, and methods used.

Studies were coded by the second author. A subsample of 10 studies was additionally coded by an experienced rater. The average interrater agreement was 91%. Core variables such as treatment type or design showed full agreement. No variable fell below 60%.

Computation of effect size

As an effect size measure, we used odds ratios (OR). This is widely recommended for dichotomous data (Fleiss 1994; Lipsey and Wilson 2001). Results in primary studies on offender treatment are usually reported as simple recidivism rates (P) or as the absolute number of successes and failures in the treatment group (TG_{Success} , TG_{Failure}) and the comparison group (CG_{Success} , CG_{Failure}), respectively. In such cases, we applied the following formulas for effect size computation:

$$OR = \frac{P_{CG} \times (1 - P_{TG})}{P_{TG} \times (1 - P_{CG})} \quad \text{and} \quad OR = \frac{CG_{\text{Failure}} \times TG_{\text{Success}}}{TG_{\text{Failure}} \times CG_{\text{Success}}}$$

If any of these frequencies equaled zero, 0.5 was added to each frequency. Some studies reported more sophisticated statistical analyses that controlled for differences between treatment and control groups. In such cases, we used these results instead of the simple recidivism rates. In logistic regression, the coefficients equal the natural log of the odds ratio (LOR), and as an exponent to e this equals the odds ratio (see Fleiss 1994). The result for the treatment variable could thus be transferred directly. In Cox regression, results are reported in the form of a rate ratio, which is similar but not identical to the odds ratio. We used the rate ratio (RR) to estimate a recidivism rate for the control group corrected for the group differences considered in the Cox regression model ($P_{CG} = RR \times P_{TG}$ or $P_{CG} = RR/P_{TG}$, depending on the coding of the treatment variable in the primary study). We then calculated the odds ratio with the above formula. Few studies reported other test statistics that could not be transformed readily into odds ratios. In these cases, we used standard procedures to calculate Cohen's d (see Lipsey and Wilson 2001) and then converted these into odds ratios using $LOR = \frac{\pi}{\sqrt{3}} \times d$ (Hasselblad and Hedges 1995, Formula 4, re-arranged) and $OR = e^{LOR}$.

If a study contained multiple dependent (sub-)samples, we used the comparison with the highest internal validity (e.g., if a study compared recidivism rates for the total sample of treated/untreated participants and additionally matched a subsample of these groups on relevant characteristics, we would use the latter comparison). Studies often reported multiple outcome variables. Different domains of recidivist behavior (i.e., sexual, violent, or general recidivism) were always analyzed separately. If a

study used different indicators of failure (i.e., charge, arrest, conviction, or lapse behavior), effect sizes were coded separately and then averaged to a single effect size.

Some studies reported separate results for different offender types or risk groups, but did not meet criteria for independent comparisons as defined above. Here, we calculated effect sizes separately for the subgroups and used the weighted average to obtain a study effect size (see Fleiss 1994).

Whenever possible, participants who dropped out of treatment were included in the treatment group ('intent to treat'). However, we evaluated the effects of treatment dropout by additionally contrasting them with both treated and untreated groups.

Integration and statistical analyses

Statistical analyses were conducted on the natural log of the odds ratio (see Fleiss 1994; Lipsey and Wilson 2001). To integrate effect sizes, we applied the weighting procedures based on the standard error of individual effect sizes (Hedges and Olkin 1985). Because of heterogeneous effect size distributions (according to the Q test of homogeneity; Hedges and Olkin 1985), we applied a random effects model. Moderator analyses were carried out under the assumption of a mixed effects model (see, also, Lipsey and Wilson 2001; Wilson 2001).

Results

Descriptive characteristics

Table 1 gives an overview of some characteristics of the studies/comparisons included for analysis. Most studies came from North America. Approximately one-third contained unpublished data. Most were relatively recent. Nearly three quarters of the studies have been published since 1990. However, as the treatment section shows, the actual program implementation started much earlier.

Nearly one-half of the comparisons addressed cognitive-behavioral programs. Due to basic similarities, we also subsumed two studies of multisystemic treatment under this category. Fourteen comparisons addressed physical therapy, eight of which dealt with surgical castration. In more than one-half of the studies, authors were affiliated with the evaluated treatment. Most treatments were specifically designed for sex offenders. However, it was rarely possible to rate whether treatment was implemented reliably. Only one-quarter of the comparisons provided a documentation of adequate program integrity. Approximately one-half of the interventions took place in an institutional setting. Although a group format was most frequent, nearly one-half of the programs included at least some individualized treatment. An explicit extension of treatment through specific aftercare services was reported for only 15 comparisons.

Table 1. Descriptive characteristics of the 80 comparisons included in the meta-analysis.

Coding variable and categories	Frequency	Percentage
<i>General study characteristics^a</i>		
Publication year		
Before 1980	7	10.1
1980s	11	15.9
1990s	29	42.0
Since 2000	22	31.9
Country		
USA	31	44.9
Canada	17	24.6
Great Britain	8	11.6
German-speaking countries	8	11.6
Other	5	7.2
Publication type		
Journal article	34	49.3
Book, chapter	10	14.5
Unpublished	25	36.2
<i>Treatment characteristics</i>		
Time of treatment implementation		
Before 1970	14	17.5
1970s	17	21.3
1980s	30	37.5
1990s	19	23.8
Mode of treatment		
Cognitive-behavioral	37	46.3
Classical behavioral	7	17.5
Insight-oriented	7	17.5
Therapeutic community	10	12.5
Other psychosocial, unclear	5	6.3
Hormonal medication	6	7.5
Surgical castration	8	10.0
Author affiliation to treatment program		
Yes	42	52.5
No	31	38.8
Unclear	7	8.8
Sex-offender-specific treatment		
Yes	64	80.0
No	9	11.3
No information available	7	8.8
Integrity of treatment implementation ^b		
Acceptable	18	25.0
Problematic	5	6.9
No information available	49	68.1
Setting of treatment		
Prison	25	31.3
Hospital	14	17.5
Outpatient	29	36.3
Mixed	10	12.5
No information available	2	2.5

Table 1. Continued.

Coding variable and categories	Frequency	Percentage
<i>Format of treatment^b</i>		
Only group treatment	18	22.5
Mainly group treatment	17	21.3
Mixed	10	12.5
Mainly individual treatment	8	10.0
Only individual treatment	7	8.8
No information available	12	15.0
<i>Aftercare</i>		
Obligatory	9	11.3
Optional	6	7.5
Not offered, not reported	65	81.3
<i>Offender characteristics</i>		
<i>Age group</i>		
Adults	45	56.3
Adolescents	7	17.5
Mixed	8	10.0
Unclear	20	25.0
<i>Homogeneity of age</i>		
High	7	17.5
Medium	23	28.8
Low	23	28.8
Unclear	27	33.8
<i>Offense type^c</i>		
Rape	44	55.0
Child molestation	59	73.7
Incest offenses	38	47.5
Exhibitionism	24	30.0
Other hands-on offenses, not specified	5	6.3
Other hands-off offenses, not specified	5	6.3
Not specified	20	25.0
<i>Treatment participation</i>		
Voluntary	37	46.3
Nonvoluntary	16	20.0
Mixed	8	10.0
Unclear	19	23.8
<i>Methodological characteristics</i>		
<i>Sample size</i>		
Up to 50	25	31.3
51–100	12	15.0
101–200	18	22.5
201–500	14	17.5
More than 500	11	13.3
<i>Maryland scale</i>		
Level 2 (nonequivalent)	48	60.0
Level 3 (equivalence assumed)	19	23.8
Level 4 (matching, statistical control)	7	8.8
Level 5 (randomization)	6	7.5

Table 1. Continued.

Coding variable and categories	Frequency	Percentage
Initial group equivalence		
Yes	20	25.0
No, TG at higher risk	14	17.5
No, CG at higher risk	4	5.0
No, direction unclear	8	10.0
No, but controlled for statistically	5	6.3
No information available	29	36.3
Control group formation		
Treatment volunteers	13	16.3
No treatment available	14	17.5
No treatment order	15	18.8
Treatment refused	19	23.8
Other	10	12.5
Unclear	9	11.3
Type of reoffense ^c		
Sexual	74	92.5
Violent	20	25.0
Any	49	61.3
Follow-up period (months)		
12–24	14	17.5
25–36	12	15.0
37–60	23	28.8
61–84	12	15.0
>84	19	23.8
Source of recidivism data		
Criminal records only	64	80.0
Self-report	6	7.5
Not indicated	10	12.5
Definition of recidivism		
Arrest	19	23.8
Conviction	24	30.0
Charge	15	18.8
Lapse behavior	3	3.8
Multiple outcomes	6	7.5
Not indicated	13	16.3

^a*n* = 69.^b*k* = 72 (no reasonable categorization for surgical castration possible).^cIndividual comparisons may cover multiple categories.

Seven comparisons referred to programs that targeted exclusively juvenile sexual offenders. Only these were very homogeneous in terms of age. Most treatment programs combined individuals with different types of sex offense. Child molestation was most frequent, followed by rape. However, no program referred exclusively to rapists. Nine programs addressed child molesters only, and four addressed exhibitionists only. Most frequently, treated offenders had participated voluntarily. Thirty percent of the comparisons referred to offenders who were at least partially obliged to attend treatment.

Sample sizes (TG + CG) varied from 15 to 2,557 ($Md = 118$). Roughly one-third of the comparisons contained less than 50 offenders. Only seven comparisons were based on a randomized design. One of these was compromised, so that only six could be coded on Level 5 of the Maryland Scale. Sixty percent of the comparisons were on Maryland Scale Level 2, that is, group equivalence could not be assumed. When group differences actually were tested and reported, the TG was more often at higher risk than the CG. However, for 29 comparisons (including all randomized trials), no information was available on group differences. In nearly one-quarter of the comparisons the CG consisted of treatment refusers.

Nearly all studies used a specific measure of sexual recidivism. Recidivism was recorded after an average follow-up period of more than 5 years (TG: $M = 63.54$ months, $SD = 42.09$; CG: $M = 62.41$, $SD = 42.37$). It was mainly based on entries in official criminal records. Few studies additionally used information from the offenders themselves. The most common definition of failure was reconviction, followed by rearrest and new charges. In three comparisons the criterion was defined rather loosely as 'inappropriate' or 'lapse' behavior. In six comparisons, outcomes were reported separately for different definitions.

Total effects

The 74 comparisons reporting data on sexual recidivism revealed an average recidivism rate of 12% for treated groups and 24% for comparison groups (unweighted average). This is a 50% reduction. However, when we calculated the recidivism rates for treated and comparison participants taking the respective sizes of TG and CG in the 74 comparisons into account (i.e., when we calculated an n -weighted average for treated and comparison groups), the difference in recidivism rates vanished completely (11% each for treated and comparison participants). These conflicting results were due to great differences in the size of TG and CG in some primary studies. Studies with very large control groups and comparatively small treated groups and an overall low recidivism base rate reduced the n -weighted average of the CG recidivism rate considerably but not the TG average. Therefore, one should not draw conclusions from these averages. Effect-size integration avoids the problem of different TG/CG sizes, because the recidivism rates are first evaluated on the level of the individual comparisons and only then integrated. The mean effect size can then be used to estimate mean CG recidivism to illustrate results. Figure 1 gives an overview of the effect sizes (logged OR) of the 74 comparisons. It shows that the majority of effects were positive ($k = 53$). Converted to OR, the effects ranged from a minimum of 0.17 to a maximum of 33.33.

We integrated the individual effect sizes according to the random model. Results are shown in Table 2. The mean OR of 1.70 for sexual recidivism was highly significant ($z = 4.96$, $P < 0.001$). The absolute difference in sexual recidivism between TG and CG was 6.4 percentage points. This is a 37% reduction

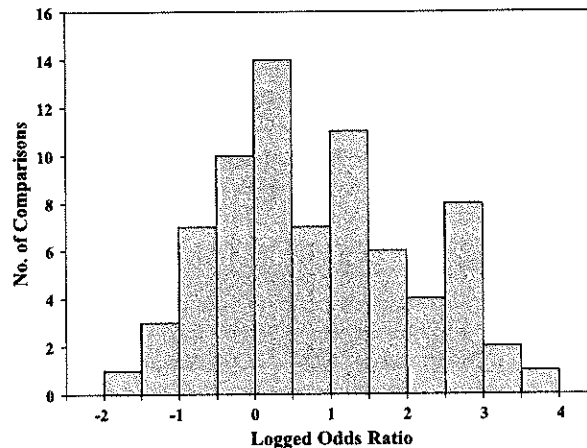


Figure 1. Distribution of effect sizes (logged odds ratios) for sexual recidivism ($k = 74$).

from the base rate of the CG. Mean effects were similar for other areas of offending. For violent recidivism, the mean OR was 1.90 ($z = 5.36$, $P < 0.001$). The average recidivism rate for treated offenders was 5.2 percentage points lower than that for untreated offenders (44% reduction). For any recidivism, the mean effect size was $OR = 1.67$ ($z = 4.52$, $P < 0.001$). The corresponding rate of general recidivism for treated offenders was 11.1 percentage points lower than for untreated offenders (31% reduction).

Except for violent recidivism, the effect size distributions showed considerable heterogeneity, $Q(73) = 237.14$, $P < 0.001$ for sexual recidivism and $Q(48) = 159.80$, $P < 0.001$ for general recidivism. Therefore, we analyzed variables that may have influenced the treatment effects. We restricted our moderator analyses to sexual recidivism as an outcome because this is the main area of interest and provides the largest database.

Table 2. Total mean effects of treatment for different areas of recidivism.

Outcome	<i>k</i>	OR	CI _{95%}	<i>Q</i>	Recidivism (%)	
					TG ^a	CG ^b
Sexual recidivism	74	1.70***	1.35–2.13	237.14***	11.1	17.5
Violent recidivism	20	1.90***	1.49–2.33	19.68	6.6	11.8
Any recidivism	49	1.67***	1.33–2.08	159.80***	22.4	32.5

k = Number of comparisons, OR = mean odds ratio, CI_{95%} = 95% confidence interval, *Q* = test of homogeneity (χ^2 , $df = k - 1$), TG = treated group; CG = comparison group.

*** $P < 0.001$.

^a*n*-Weighted average.

^bEstimated recidivism rate.

Moderator analyses

For reasons of space, we only report a selection of moderator analyses. Table 3 reports the results on variables that revealed significant differences or are particularly relevant for treatment practice. In all analyses, we have to take into account that moderating effects may be confounded with the influence of other variables.

Treatment characteristics. The various treatment approaches differed considerably in effect size. In total, physical treatment had higher effects ($OR = 7.37$, $CI_{95\%}: 4.14-13.11$, $z = 6.80$, $P < 0.001$) than did nonphysical (psychosocial) interventions ($OR = 1.32$, $CI_{95\%}: 1.07-1.62$, $z = 2.60$, $P = 0.01$), $Q(1, k = 66) = 30.47$, $P < 0.001$. This was particularly due to the very large mean ES for surgical castration. However, hormonal treatment also showed a higher effect than any of the psychosocial measures. Of these, only cognitive-behavioral treatments and classic behavior therapy had a significant impact on sexual recidivism. With OR close to 1, the other approaches did not influence recidivism significantly. A regression analysis entering the more differentiated 4-point scale ratings on the use of individual treatment elements confirmed these findings. It showed significant standardized regression weights (β) for a cognitive orientation (0.36), behavioral conditioning techniques (0.26), and hormonal medication (0.26); all $P < 0.05$.

In the further moderator analyses, we excluded the studies on surgical castration for the following reasons: (a) This approach differs strongly from the others. (b) It is currently rarely used in practice. (c) The effect sizes for castration were extremely large and homogeneous, $Q(7) = 1.32$, $P = 0.99$, and thus would have unduly biased the results.

For the remaining 66 comparisons, the mean effect remained significant ($OR = 1.38$, $CI_{95\%}: 1.13-1.69$, $z = 3.16$, $P < 0.01$). A test of homogeneity still indicated a highly heterogeneous effect size distribution, $Q(65) = 163.92$, $P < 0.001$, justifying further moderator analyses.

As Table 3 shows, the decade in which the program was implemented related significantly to effect size. However, there was no linear relationship. This was also confirmed in a correlation analysis ($r = 0.16$, $P = 0.15$). Thus, more 'modern' programs did not generally prove to be particularly successful. The year of publication as another indicator of recency showed a similar picture. Here, the differences were even less pronounced ($r = 0.08$, $P = 0.51$). When the study authors had been involved in the treatment program, there was a larger effect. When there was no such affiliation, the respective OR was close to 1.

Only programs designed specifically for sex offenders had a significant effect. The few others even showed a negative outcome. Although the setting variable revealed no significant difference, there was a strong tendency for relatively larger effects in outpatient treatment and smaller effects in institutions. Mixed settings had an intermediate ES. A linear order from institutional to outpatient treatment showed a significant correlation ($r = 0.27$, $P = 0.02$).

Whether the treatment was delivered in an individual or a group format did not result in significant outcome differences. However, in this category, we must assume

Table 3. Moderator analyses.

Variables	k	Q_{het}	OR	CI _{95%} Lower-upper
<i>Treatment characteristics</i>				
Treatment approach		36.02***		
Cognitive-behavioral	35		1.45**	1.12–1.86
Classic behavioral	7		2.19**	1.22–3.92
Insight oriented	5		0.98	0.51–1.89
Therapeutic community	8		0.86	0.54–1.35
Other psychosocial, unclear	5		0.94	0.53–1.65
Hormonal medication	6		3.08**	1.40–6.79
Surgical castration	8		15.34***	7.34–32.05
Time of treatment implementation		7.37**		
Before 1970	5		0.56*	0.32–0.98
1970s	14		2.03***	1.34–3.09
1980s	30		1.38**	1.08–1.77
1990s	17		1.27	0.86–1.87
Specific treatment sex offenders		4.70*		
Yes	56		1.56***	1.27–1.93
No	5		0.76	0.41–1.41
Setting of treatment		5.10		
Prison	21		1.16	0.84–1.60
Hospital	8		1.10	0.62–1.94
Outpatient	27		1.93***	1.35–2.77
Mixed	10		1.37	0.78–2.41
Format of treatment		6.74		
Only group treatment	17		1.12	0.76–1.66
Mainly group treatment	17		1.57*	1.02–2.42
Mixed	8		2.45*	1.36–4.40
Mainly individual treatment	8		1.40	0.77–2.53
Only individual treatment	6		2.88*	1.14–7.24
Author affiliation to treatment program		10.95***		
Yes	32		1.92***	1.44–2.56
No	30		0.99	0.76–1.29
<i>Offender characteristics</i>				
Age group		1.19		
Adolescents only	7		2.35*	1.01–5.43
Adults only	36		1.43*	1.08–1.90
Offense type		9.04*		
Rape	5		4.91**	1.64–14.68
Child molestation (extrafam.)	9		2.15*	1.11–4.16
Child molestation (incl. incest)	10		1.02	0.58–1.80
Exhibitionism	4		3.72*	1.27–10.93
Treatment participation		2.22		
Voluntary	28		1.45*	1.08–1.93
Nonvoluntary	15		1.05	0.70–1.58
Mixed	7		1.01	0.57–1.77
Treatment termination		— ^a		
Treatment completed regularly	44		1.58***	1.23–2.05
Dropped out of treatment	14		0.51***	0.39–0.67

Table 3. Continued

Variables	<i>k</i>	<i>Q_{bet}</i>	OR	CI _{95%} Lower-upper
<i>Methodological characteristics</i>				
Sample size		31.43***		
Up to 50	18		4.03***	2.50-6.50
51-100	10		1.32	0.76-2.27
101-200	16		1.65**	1.13-2.41
201-500	12		1.00	0.72-1.38
More than 500	10		0.88	0.64-1.21
Maryland Scale		6.13		
Level 2 (nonequivalent)	37		1.16	0.90-1.50
Level 3 (equivalence assumed)	17		2.08***	1.40-3.08
Level 4 (matching, statistical control)	6		1.19	0.67-2.12
Level 5 (randomization)	6		1.48	0.74-2.96
Control group formation		1.64		
Treatment refused	11		1.96**	1.20-3.20
Other	47		1.37*	1.07-1.75
Source of recidivism data		4.56*		
Criminal records only	57		1.28*	1.04-1.57
Also self-report	5		3.32**	1.42-7.78

Except for the analysis on treatment approaches, studies on surgical castration are not included in the moderator analyses.

k = number of comparisons; *Q_{bet}* = test of between group differences (χ^2 -distributed with *df* = number of categories - 1); OR = odds ratio; CI_{95%} = 95% confidence interval; CG = comparison group.

*Comparisons are based on identical CG in part; between-group differences could thus not be tested statistically.

**P* < 0.05.

***P* < 0.01.

****P* < 0.001.

confounding with various content variables. For example, both hormonal medication and systemic treatment had to be subsumed under the individual category.

Offender characteristics. Programs that specifically addressed juvenile sex offenders had a higher effect than those for adult offenders. However, this difference was not significant. A related analysis showed that treatment of age-homogeneous groups tended to be more successful (*k* = 48; *r* = 0.23, *P* = 0.10).

Although the impact on specific offender groups is highly important in treatment practice, only few studies differentiated offense categories. These comparisons showed significant effects for all categories except that of intra-familial child molesting. The latter finding is related to the low recidivism base rate for incest offenders. There was a relatively large effect for rapists, but this was based on only five studies.

When sexual offenders participated voluntarily in treatment, the average *ES* was significantly positive. Obligatory participation and mixed conditions resulted in no effect. However, these differences were not significant.

Whether treatment was terminated regularly or prematurely had an impact on sexual recidivism. Whereas regular completers showed better effects than the control groups, dropouts did significantly worse. Dropping out of treatment doubled the odds of relapse and this negative effect was even homogeneous, $Q(13) = 11.52$, $P = 0.57$. In contrast, effect sizes that referred to completers revealed considerable heterogeneity, $Q(43) = 100.20$, $P < 0.001$.

Methodological characteristics. Sample size correlated significantly with effect size ($r = -0.26$, $P = 0.03$). This was particularly due to the extreme poles of the sample sizes. Comparisons based on small samples ($N \leq 50$) showed very clear effects, whereas the mean OR for very large samples was slightly below 1. This relationship could not be attributed to a publication bias only: Although in unpublished studies, the effect was somewhat less pronounced at the extreme ends of the sample size distribution, there was an even clearer linear trend compared with published studies ($r = -0.34$ vs. $r = -0.20$).

Overall, design quality did not yield a significant moderator effect. Comparisons of equivalent TG and CG (Maryland Scale Level 3 and above) revealed an average OR of 1.69 (CI_{95%}: 1.26–2.28). At $P = 0.06$, this exceeded the OR of 1.16 for Level 2 comparisons. However, as Table 3 shows, there was no linear relationship between design quality and ES. Randomized trials also did not differ from the other comparisons, $Q(1, k = 66) = 0.07$, $P = 0.79$. Control groups containing treatment refusers revealed relatively large effects, however, these effects did not differ significantly from studies using other control groups.

The length of follow-up did not correlate with ES ($r = 0.00$). Different indicators of reoffending (i.e., reconviction, rearrest, etc.) also did not relate systematically to outcome variation, $Q(6, k = 60) = 3.45$, $P = 0.49$. In contrast, the sources used to gather the respective information had a significant impact on ES, $Q(2, k = 62) = 7.91$, $P = 0.02$. Comparisons using not only official records but also self-reported data had larger effects. However, this variable was confounded with the type of treatment, because all studies on hormonal medication included self-reported recidivism.

As could be expected, a higher baserate of recidivism correlated with a larger ES ($r = 0.30$, $P = 0.01$). This effect was also confounded, because informal data sources produced higher baserates. After controlling for this aspect, the relationship between baserate and ES was weaker ($\beta = 0.23$, $P = 0.08$).

Features of descriptive validity do not address the process of treatment and its evaluation but primarily the process of scientific reporting. Nonetheless, our analysis showed that these also related to effect size. Both the quality of documenting the treatment concept ($r = 0.33$, $P < 0.01$) and the reporting of outcome statistics ($r = 0.24$, $P = 0.03$) correlated significantly with ES.

General study characteristics. There were no significant ES differences between the various groups of countries in which the studies were performed, $Q(4, k = 66) = 2.46$, $P = 0.65$. Regarding publication type, we only found a significant effect for published comparisons ($k = 40$, OR = 1.62, CI_{95%}: 1.23–2.13, $P < 0.001$). The

mean effect for unpublished comparisons was only $OR = 1.14$ ($k = 26$, $CI_{95\%} = 0.84-1.54$, $P = 0.42$). However, this difference was not significant, $Q(1, k = 66) = 2.91$, $P = 0.09$.

Sensitivity analysis. The effects of moderators may be influenced strongly by a few results from studies with very large sample size (Lipsey and Wilson 2001). Therefore, we conducted a sensitivity analysis by using two different procedures. First, we excluded all comparisons with sample sizes larger than 1,000 ($k = 3$). In a second approach, all comparisons with a sample size of more than 500 were truncated to $n = 500$. With one exception, the sensitivity analyses confirmed the significant moderator effects reported in Table 3. Only the effect of unspecific offender treatment failed to reach significance when we eliminated the comparisons with sample sizes larger than 1,000 ($P = 0.08$).

Hierarchical regression. The previous analyses have repeatedly indicated problems of confounded moderators. Therefore, it is particularly relevant to see how far treatment effects are confounded with methodological and other characteristics of the evaluation. To answer this question, we computed a hierarchical regression analysis controlling sequentially for those proportions of outcome variance that could not be attributed to the treatment itself. At first, we entered unspecific and methodological study characteristics into the model. We then added offender characteristics, general treatment characteristics, and, finally, the treatment content. We entered variables that were theoretically important or empirically significant on the bivariate level ($r \geq 0.20$). At each hierarchical step, variables that did not contribute to the explanation of variance were excluded stepwise ($P > 0.10$). Because only a relatively small number of comparisons were available for the analysis, we chose this procedure in order to not overload the model with insignificant variables. Missing values were plugged with the sample mean, and analyses controlled for the effects of missing values (see Cohen and Cohen 1983). In contrast to the previous bivariate analyses, the hierarchical regression was based on a fixed effects model because we expected that the variables included would reduce a considerable part of the observed heterogeneity. The results are presented in Table 4.

As expected, heterogeneity was of moderate magnitude and not significant, $Q(55, k = 66) = 65.40$, $P = 0.16$. With 60%, the model explained a large proportion of ES variance, $Q(10, k = 66) = 98.52$, $P < 0.001$. However, one should bear in mind that due to the stepwise exclusion of variables on each cluster level, the model becomes artificially 'clean.' Only a few variables remained as independent predictors in each cluster. Methodological characteristics accounted for a considerable amount of variance (45%). Due to deficits in sample description and differentiation, offender characteristics only had a small independent impact on ES. Age homogeneity was the only relevant variable, and added 3% of explained variance. General characteristics of treatment were more important. Specificity of treatment for sex offenders, involvement of authors in the program, and a group format contributed to a 9 percentage points increase in

Table 4. Hierarchical regression.

Variable cluster	ΔR^2
Methodological characteristics	0.45***
+ Quality of outcome reporting (0.35***), Quality of treatment description (0.20*), Small sample, $N \leq 50$ (0.42***), Treatment refusers as CG (0.16 [†])	
- TG contains dropouts (-0.24**)	
Offenders	0.03*
+ Age homogeneity of TG (0.16*)	
General treatment characteristics	0.10***
+ Involvement of authors (0.24**), Group format (0.18 [†])	
- Not specific for sexual offenders (-0.19 [†])	
Content of treatment	0.03*
+ Cognitive orientation (0.28*)	

Changes in index direction correspond to higher (+) and lower (-) effect sizes respectively (standardized β weights are reported in brackets). Total $R^2 = 0.60$, $Q(10) = 98.52$, $P < 0.001$.

[†] $P < 0.10$.

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

explained ES variance. Although it was only entered in the final step, a cognitive treatment orientation still added significantly to the explanation of ES variance over and above the preceding clusters. None of the other treatment variables remained in the model. Obviously, the effects of hormonal medication and behavioral conditioning methods were highly confounded with other variables. Taken together, the last two steps suggest that at least one fifth of the explained ES variance could be attributed to treatment characteristics. When interpreting this figure, we should bear in mind that this is a very conservative estimate because all other variables had been controlled already.

Discussion

Due to a recent increase in research and the multilingual approach of our review, this meta-analysis contains 80 comparisons between treatment and control groups containing a total of more than 22,000 individuals. This is currently the most comprehensive database on the outcome of sex offender treatment. Nearly one-third of the studies have been published since 2000, and approximately one-third come from countries outside of North America. These are indicators of a strong international interest in 'what works' for sex offenders. However, even though we have excluded studies containing no control group or only a comparison with dropouts, the methodological quality of the studies still remains moderate. Only 40% of the comparisons reach a level of 3 or higher on the Maryland Scale of Methodological Rigor (Sherman et al. 1997), indicating sufficient control of equivalence between TG and CG. Only seven evaluations contain a randomized

design. We have not restricted our review to these 'gold-standard' studies for the following reasons: First, limitation to a few evaluations of heterogeneous modes of treatment would not allow a differentiated analysis. Second, even randomization does not guarantee full equivalence of TG and CG (see, e.g., Marques et al. 2005). Third, as far as nonequivalence can be assessed, it tends to promote a conservative estimate of treatment efficacy (more high-risk cases in the TGs). And last but not least, the effects of our subsample of randomized studies do not differ significantly from evaluations with lower design quality (see, also, Lipsey and Wilson 1998; Lösel 1995).

Bearing the methodological problems in mind, one should draw very cautious conclusions from our meta-analysis. The most important message is an overall positive and significant effect of sex offender treatment. The mean odds ratio is 1.70 for sexual recidivism. The equivalent d coefficient of 0.29 lies within the typical range found in meta-analyses of general offender treatment (Lösel 1995; McGuire 2002). Sexual offender treatment also has an effect on general recidivism (OR = 1.67). Obviously, effective programs do not just influence sexually motivated problem behavior but also have a broader impact on criminality. This is in accordance with the experience that many sex offenders are not 'specialized' but engage in nonsexual offenses as well (Hanson and Bussière 1998). However, our analysis also shows that unspecific offender programs have no impact on sexual recidivism.

The mean rate of sexual recidivism is 11.1% in TGs and 17.5% in CGs. At first glance, this absolute difference of a little more than 6 percentage points may seem small. However, when the low base rate of sexual recidivism is taken into account, this is equivalent to a reduction of nearly 37%. For general recidivism, the reduction is 31%. Particularly in sexual recidivism, our general effect is larger than that found by Hanson et al. (2002) in their meta-analysis of psychological treatment (27%). Most probably, this is due to our inclusion of both psychological and medical modes of treatment. The average effect of physical treatment is much larger than that of psychosocial programs. The main source for this difference is a very strong effect of surgical castration, although hormonal medication also shows a relatively good outcome.

Although the very large effect of surgical castration seems to be well replicated (eight comparisons with homogeneous ES), it calls for further comment: None of the castration studies attain Level 3 on the Maryland Scale. Accordingly, we cannot assume equivalence between the TG and CG in these evaluations. Sex offenders receiving surgical castration are a highly selected and motivated group. They apply for this very intensive intervention voluntarily, whereas control individuals often refuse it or are not accepted by expert committees (e.g., Wille and Beier 1989). Hence, the TGs probably are at lower risk of reoffending than the CGs. For ethical, legal, and medical reasons, surgical castration is also rarely used in practice (Rösler and Witztum 2000). However, the very low rate of sexual recidivism in castrated offenders suggests that societies should not abandon this approach right away but perform a differentiated assessment of the pros and cons. Within an impartial and thorough process of informed consent, it may be an option

for a subgroup of high-risk sex offenders who otherwise would receive very long or lifetime detention.

Most sex offenders do not have an abnormally high level of male sex hormones (Hucker and Bain 1990; Fedoroff and Moran 1997). As with surgical castration, we must take this into account when considering the relatively strong effect of hormonal medication. Treatment with medroxy-progesteroneacetate (in the US; e.g., Provera) or cyproteroneacetate (in Europe; e.g., Androcur) does not seem to work by normalizing extreme testosterone levels, but by strongly reducing more or less normal levels of sexual arousal (Rösler and Witztum 2000). In addition, there are serious negative side effects that frequently lead to noncompliance and dropout (e.g., Langevin 1979). The termination of medication may rapidly increase the risk of recidivism (Meyer et al., 1992). Therefore, hormonal medication is indicated primarily for cases in which sexual arousal plays a central role in offending (e.g., Hall 1996) and should be accompanied by psychological treatment that supports compliance and has its own causal effect on sexual reoffending (Maletzky 1991; Meyer and Cole 1997).

This is why the pharmacological studies in our meta-analysis often contain psychosocial interventions as well. We have analyzed the impact of such 'treatment packages' by rating the various components separately. A regression analysis shows that only three modes of treatment have a significant impact: hormonal, behavioral, and cognitive-behavioral. The results of the first two types of program are more confounded with methodological and other study characteristics than those of the latter. After controlling for such variables, only the cognitive-behavioral orientation shows an independent treatment effect (see Table 4).

The significant positive effect of cognitive-behavioral programs is based on a solid number of 35 independent comparisons. With seven comparisons, the significant effect of classic behavior therapy has a much smaller database. The same applies to insight-oriented treatment, therapeutic communities, and other types of psychosocial programs that reveal no significant effect. That well-structured cognitive-behavioral programs work relatively well is in accordance with the literature on general offender treatment (e.g., Lösel 2001a; McGuire 2002). It is also consistent with the findings of previous reviews of sex offender treatment (e.g., Gallagher et al. 2000; Hall 1995; Hanson et al. 2002). However, the effect size for cognitive-behavioral programs in our analysis ($OR = 1.45$) is slightly smaller than that reported by Hanson et al. for 'current' programs that consist of mainly cognitive-behavioral approaches ($OR = 1.67$; direction converted by us).

Overall, we have not found that more recent programs are superior in outcome. Although treatment before the 1970s was clearly ineffective, neither programs from the 1990s nor publications after 2000 reveal stronger effects than in previous decades. Even within the cognitive-behavioral category, more current programs are not more effective than older ones. Some recent evaluations have revealed rather small or no positive effects (e.g., Friendship et al. 2003; Hanson et al. 2004; Ruddiis and Timmerman 2000; Worling and Curwen 2000). A follow-up of one of the soundest evaluations has also found no positive effect (Marques et al. 2005).

Due to the necessary follow-up lags, even recent studies may not represent all the features of the current state of the art in sexual offender treatment. One must also bear in mind that outcomes of treatment often decline when model projects are transformed into routine practice (Lösel 2001b).

The heterogeneity of outcomes within similar types of programs may be partially explained by the impact of other factors on effect size. Similar to the treatment content, these further moderators must be interpreted very cautiously: (a) Some effects are based on only a few studies. (b) The random model is less sensitive for moderator effects (Overton 1998). (c) Multiple significance testing in moderator analyses enhances the risk of an alpha error. (d) The moderators are confounded and some have no impact on the multivariate level.

Although we have found no linear relationship between design quality and outcome, there is a tendency of larger effects in studies containing equivalent treatment and control groups (at least Level 3 on the Maryland Scale). More randomized studies on the same types of treatment are needed to clarify this issue in the field of sexual offender treatment in a similar way as in other criminological areas (e.g., Weisburd et al. 2001). The impact of other methodological characteristics appears more clearly. For example, studies that include not only official recidivism data but also self-reports show larger effects. This finding is partially confounded with treatment by hormonal medication. Issues of descriptive validity such as quality of treatment description and outcome reporting are also related to larger effects.

In practical terms, the relation between sample size and treatment effectiveness is particularly important. Small studies ($N \leq 50$) reveal a large ES and large studies ($N > 500$) a small ES. One explanation of this result relates to publication bias. Larger samples are more likely to reveal the significance of a true small effect (Weisburd et al. 2003). Due to author or editor decisions, such large studies may be published, whereas small studies, which would have needed a larger effect size to attain significance, remain unpublished. In accordance with such an interpretation, published studies have a larger effect than unpublished studies. However, this difference is not significant, and we have found a similar – even somewhat greater – impact of sample size among the unpublished studies. Of course, this does not fully rule out some kind of publication bias, because negative results may also be less likely to be reported in unpublished studies (particularly when the researcher has a strong vested interest in the success of the program). Nevertheless, we must take a second explanation into account: In large samples, it is more difficult to maintain integrity and homogeneity of treatments or samples, and this is related to the effect size itself (Lösel and Wittmann 1989; Weisburd et al. 1993). A further finding supports the interpretation in terms of integrity: Programs in which the study authors were involved have a larger effect. As most outcome measures are beyond the influence of authors, this finding may indicate a more thorough implementation and monitoring of the program. Similar results on the effect of small samples and/or author involvement have been observed in juvenile offender treatment (Lipsey and Wilson 1998) and developmental prevention of antisocial behavior (Farrington and Welsh 2003; Lösel and Beelmann 2003). The few comparisons for which treatment integrity can actually be rated provide further support for this relationship.

The context of treatment is also relevant for outcome. Ambulatory programs have larger effects than institutional treatment. Because we have analyzed only control group evaluations, this finding cannot be reduced to a different risk in the respective offender groups. Offender characteristics also have some impact on effect size. However, there is often a lack of detailed information on the samples. In nearly half of the studies, samples cannot even be differentiated according to the type of sexual offense. As far as this was possible, we found a significant difference in outcome. This is mainly due to a zero effect on incest child molesters resulting from the very low base rate of (official) recidivism in this group. More treatment evaluations on specific subgroups of offenders are needed to form a solid empirical basis for differential indication.

Voluntary treatment leads to a slightly better outcome than mandatory participation, and programs for adolescents are a little more effective than those for adults. Although, these differences are not significant and may be confounded with a larger base rate of recidivism in juveniles, age homogeneity is a significant moderator in the hierarchical regression. A more pronounced finding is the higher recidivism among treatment dropouts. As a consequence, studies that include dropouts in the treatment group have smaller effects. The high risk of recidivism in dropouts underlines that this group is a core problem in offender rehabilitation and controlled evaluation (Lösel 2001b). It should not only be interpreted as an individual deficit of the offender but as an interactive process and lack of fit between the program and the offender's needs and motivations (McMurrin 2002). Systematic processes of program accreditation and quality management like those in Canada, England and Wales, or Scotland may help to reduce this and other problems in offender treatment. However, from a realistic perspective, we should not expect too much within a short time.

Overall, there is evidence for a positive effect of sexual offender treatment. Cognitive-behavioral and hormonal treatment are most promising. In addition, various other moderators are related to a better or worse outcome. In particular, methodological factors play an important role and seem to be confounded with treatment and offender characteristics. This problem of confounded moderators is rather general and difficult to solve (Lipsey 2003). Our hierarchical regression is only a first attempt to disentangle such patterns in the field of sexual offender treatment. We need more high-quality outcome studies that address specific subgroups of sex offenders as well as more detailed process evaluations on various treatment characteristics and components. Implementing such strategies in research and practice will further clarify 'What works for whom under which circumstances?'

Acknowledgements

We wish to thank all our colleagues who provided us with information on unpublished studies on sexual offender treatment.

Appendix: Studies integrated into the meta-analysis

- Allam, J. (1998). *Effective practice in work with sex offenders: A reconviction study comparing treated and untreated offenders*. Birmingham: West Midlands Probation Service Sex Offender Unit.
- Bakker, L., Hudson, S. M., Wales, D. S. & Riley, D. (1998). "And there was Light": *Evaluating the Kia Marama Treatment Programme for New Zealand sex offenders against children*. Christchurch, NZ: Psychological Service, Department of Corrections.
- Barbaree, H. E. & Seto, M. C. (1998). *The ongoing follow-up of sex offenders treated at the Warkworth Sexual Behaviour Clinic*. Toronto: Centre for Addiction and Mental Health.
- Barnes, J. M. (2000). *Recidivism in sex offenders: A follow-up comparison of treated and untreated sex offenders released to the community in Kentucky*. Unpublished doctoral dissertation, University of Louisville, Louisville, KY.
- Berner, W. & Karlick-Bolten, E. (1986). *Verlaufsformen der Sexualkriminalität. 5-Jahres-Katamnesen bei 326 Sexualdelinquenten unter Berücksichtigung von Frühsozialisation, vorausgegangener Delinquenz, psychiatrisch-psychologischer Diagnostik und Therapie* [Developmental courses of sexual criminality]. Stuttgart: Enke.
- Bluglass, R. (1980). Indecent exposure in the West Midlands. In D. West (Ed.), *Sex offenders in the criminal justice system* (pp. 171–180). Cambridge: Cambridge Institute of Criminology.
- Borduin, C. M., Henggeler, S. W., Blaske, D. M. & Stein, R. J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology* 34, 105–113.
- Borduin, C. M. & Schaeffer, C. M. (2001). Multisystemic treatment of juvenile sexual offenders: A progress report. *Journal of Psychology and Human Sexuality* 13, 25–42.
- Byrne, S. M. (1999). *Treatment efficacy of a juvenile sexual offender treatment program*. Unpublished doctoral dissertation, Memorial University of Newfoundland, St. Johns, Newfoundland, Canada.
- Cornu, F. (1973). *Katamnesen bei kastrierten Sittlichkeitsdelinquenten aus forensisch-psychiatrischer Sicht* [Follow-ups with castrated sex offenders]. Basel: Karger.
- Craissati, J., Falla, S., McClurg, G. & Beech, A. (2002). Risk, reconviction rates and pro-offending attitudes for child molesters in a complete geographical area of London. *Journal of Sexual Aggression* 8, 22–38.
- Di Fazio, R., Abracen, J. & Looman, J. (2001). Group versus individual treatment of sex offenders: A comparison. *Forum on Corrections Research* 13, 56–59.
- Düinkel, F. & Geng, B. (1994). Rückfall und Bewährung von Karrieretätern nach Entlassung aus dem sozialtherapeutischen Behandlungsvollzug und dem Regelvollzug [Relapse and non-relapse of career offenders after release from socialtherapeutic and standard prisons]. In M. Steller, K.-P. Dahle & M. Basqué (Eds.), *Straftäterbehandlung: Argumente für eine Revitalisierung in Forschung und Praxis* (pp. 35–74). Pfaffenweiler: Centaurus.
- Fedoroff, J. P., Wisner-Carlson, R., Dean, S. & Berlin, F. S. (1992). Medroxy-progesterone acetate in the treatment of paraphilic sexual disorders: Rate of relapse in paraphilic men treated in long-term group psychotherapy with or without medroxy-progesterone acetate. *Journal of Offender Rehabilitation* 18, 109–123.
- Friendship, C., Mann, R. E. & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence* 18, 744–759.

- Frisbie, L. V. (1969). *Another look at sex offenders in California* (Mental Health Research Monograph No. 12). Sacramento, CA: State of California, Department of Mental Hygiene.
- Groth, N. A. (1983). Treatment of the sexual offender in a correctional institution. In J. G. Greer & I. R. Stewart (Eds.), *The sexual aggressor* (pp. 160–176). New York: Van Nostrand-Reinhold.
- Guarino-Ghezzi, S. & Kimball, L. M. (1998). Juvenile sex offenders in treatment. *Corrections Management Quarterly* 2, 45–54.
- Hall, G. C. N. (1995). The preliminary development of a theory-based community treatment for sexual offenders. *Professional Psychology: Research and Practice* 26, 478–483.
- Hanson, R. K., Steffy, R. A. & Gauthier, R. (1992). Long-term follow-up of child molesters: Risk predictors and treatment outcome.
- Hanson, R. K. & Nicholaichuk, T. P. (2000). A cautionary note regarding Nicholaichuk et al. (2000). *Sexual Abuse: A Journal of Research and Treatment* 12, 289–293.
- Hanson, R. K., Broom, I. & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Sciences* 36, 87–96.
- Hedderman, C. & Sugg, D. (1996). *Does treating sex offenders reduce offending?* (Research Findings No. 45). London: Home Office Research and Statistics Directorate.
- Huot, S. J. (2002). *Recidivism, recidivism, recidivism! An update of several Minnesota recidivism studies*. Paper presented at the 21st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers on 'Best Practices: Clinical and Research Collaborations,' Montréal, Québec, Canada, October.
- Janner, J. (1959). Über die Bedeutung psychiatrischer Betreuung verurteilter Sexualdelinquenten [About the importance of psychiatric care for convicted sexual offenders]. *Schweizerische Zeitschrift für Strafrecht* 74, 310–319.
- Kansas Department of Corrections. (2000). *Offender programs evaluation: Volume IV*. Topeka: Kansas Department of Corrections.
- Kaul, J., Huot, S. J., Epperson, D. & Dornfeld, M. (1994). *Sex offenders released in 1988* (Unpublished report). St. Paul, MN: Minnesota Department of Corrections.
- Lab, S. P., Shields, G. & Schondel, C. (1993). Research note: An evaluation of juvenile sexual offender treatment. *Crime and Delinquency* 39, 543–553.
- Langevin, R., Paitich, D., Hucker, S. J., Newman, S., Ramsay, G. & Pope, S., et al. (1979). The effect of assertiveness training, Provera, and sex of therapist in the treatment of genital exhibitionism. *Journal of Behavior Therapy and Experimental Psychiatry* 10, 275–282.
- Looman, J., Abracen, J. & Nicholaichuk, T. P. (2000). Recidivism among treated sexual offenders and matched controls: Data from the Regional Treatment Centre (Ontario). *Journal of Interpersonal Violence* 15, 279–290.
- Lowden, K., Hetz, N., Harrison, L., Patrick, D., English, K. & Pasini-Hill, D. (2003). *Evaluation of Colorado's prison therapeutic community for sex offenders: A report of findings*. Denver, CO: Office of Research and Statistics, Division of Criminal Justice.
- Maletzky, B. M. (1991). The use of medroxyprogesterone acetate to assist in the treatment of sexual offenders. *Annals of Sex Research* 4, 117–129.
- Maletzky, B. M. & Field, G. (2003). The biological treatment of dangerous sexual offenders. A review and preliminary report of the Oregon pilot depo-Provera program. *Aggression and Violent Behavior* 8, 391–412.
- Marques, J. K. (1999). How to answer the question, "Does sex offender treatment work?" *Journal of Interpersonal Violence* 14, 437–451.

- Marshall, W. L. & Barbaree, H. E. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behaviour Research and Therapy* 26, 499–511.
- Marshall, W. L., Eccles, A. & Barbaree, H. E. (1991). The treatment of exhibitionists: A focus on sexual deviance versus cognitive and relationship features. *Behaviour Research and Therapy* 26, 129–135.
- McConaghy, N., Blaszczyński, A. & Kidson, W. (1988). Treatment of sex offenders with imaginal desensitization and/or medroxyprogesterone. *Acta Psychiatrica Scandinavica* 77, 199–206.
- McGrath, R. J., Hoke, S. E. & Vojtisek, J. E. (1998). Cognitive-behavioral treatment of sex offenders. A treatment comparison and long-term follow-up study. *Criminal Justice and Behavior* 25, 203–225.
- McGrath, R. J., Cumming, G. F., Livingston, J. A. & Hoke, S. E. (2003). Outcome of a treatment program for adult sex offenders. From prison to community. *Journal of Interpersonal Violence* 18, 3–17.
- McGuire, T. J. (2000). Correctional institution based sex offender treatment: A lapse behavior study. *Behavioral Sciences and the Law* 18, 57–71.
- Meyer, W. J., Cole, C. & Emory, E. (1992). Depo provera treatment for sex offending behavior: An evaluation of outcome. *Bulletin of the American Academy of Psychiatry and the Law* 20, 249–259.
- Mulloy, R. & Smiley, W. C. (1996). *Recidivism and treated sexual offenders*. Paper presented at the International Congress of Psychology, August 16–21, Montréal, Canada.
- Nicholaichuk, T. P. (1996). Sex offender treatment priority: An illustration of the risk/need principle. *Forum on Corrections Research* 8, 38–41.
- Nutbrown, V. & Stasiak, E. (1987). Research monograph: A retrospective analysis of O.C.I. cost effectiveness 1977–1981. *Ontario Correctional Institute* 2, 1–16.
- Ohio Department of Rehabilitation and Correction. (1996). *Five year recidivism follow-up of sex offender releases* (Unpublished Report). Columbus, OH: Ohio Department of Rehabilitation and Correction.
- Oregon Department of Corrections. (1997). *Outcome evaluation of the Jackson County Sex Offender Supervision and Treatment Program*. Retrieved 5/29/01, from <http://www.doc.state.or.us/research/jackrpt2.pdf>
- Ortmann, R. (2002). *Sozialtherapie im Strafvollzug: Eine experimentelle Längsschnittstudie zu den Wirkungen von Strafvollzugsmaßnahmen auf Legal- und Sozialbewährung* [Social therapy in prison]. Freiburg: Edition iuscrim.
- Perkins, D. (1987). A psychological treatment programme for sex offenders. In B. J. MacGurk & D. M. Thornton (Eds.), *Applying psychology to imprisonment: Theory and practice* (pp. 191–217). London: Her Majesty's Stationery Office Books.
- Peters, J. J., Pedigo, J., Steg, J. & McKenna, J., Jr. (1968). Group psychotherapy of the sex offender. *Federal Probation* 32, 41–45.
- Procter, E. (1996). A five-year outcome evaluation of a community-based treatment program for convicted sexual offenders run by the probation service. *Journal of Sexual Aggression* 2, 3–16.
- Rasmussen, L. A. (1995). *Factors related to recidivism among juvenile sexual offenders*. Unpublished doctoral dissertation, University of Utah.
- Rattenbury, F. R. (1985). *The outcomes of hospitalized and incarcerated sex offenders: A study of offender types, recidivism rates, and identifying characteristics of the repeat offender*. Unpublished doctoral dissertation, Loyola University of Chicago, Chicago, IL.
- Rice, M. E., Harris, G. T. & Quinsey, V. L. (1993). Evaluating treatment programs for child

- molesters. In J. Hudson & J. V. Roberts (Eds.), *Evaluating justice: Canadian policies and programs* (pp. 189–203). Toronto: Thompson.
- Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders* (Research Report No. R-41). Ottawa, Ontario, Canada: Correctional Service of Canada.
- Romero, J. J. & Williams, L. M. (1983). Group psychotherapy and intensive probation supervision with sex offenders: A comparative study. *Federal Probation* 47, 36–42.
- Ruddijs, F. & Timmerman, H. (2000). The Stichting ambulante preventie projecten method: A comparative study of recidivism in first offenders in a Dutch outpatient setting. *International Journal of Offender Therapy and Comparative Criminology* 44, 725–739.
- Schmid, P. (1988). *Was geschieht mit den Sexualstraftätern in der Psychiatrie? Darstellung und Bewertung der psychiatrischen Behandlung von Sexualstraftätern im Psychiatrischen Landeskrankenhaus Bad Schussenried in den Jahren 1978–1987* [What happens to sexual offenders in psychiatric clinics]. Unpublished doctoral dissertation, Universität Tübingen, Tübingen.
- Smith, M. A. (2000). *The impact of prison-based drug treatment on the timing and risk of reincarceration of sexual offenders*. Unpublished dissertation, Florida State University, Tallahassee, FL.
- Song, L. & Lieb, R. (1995). *Washington State sex offenders: Overview of recidivism studies* (No. 95-02-1101). Olympia, WA: Washington State Institute for Public Policy.
- Stalans, L. J., Seng, M., Yarnold, P., Lavery, T. & Swartz, J. (2001). *Process and initial impact evaluation of the Cook County Adult Probation Department's Sex Offender Program: Final and summary report for the period of June, 1997 to June, 2000*. Chicago, IL: Illinois Department of Corrections.
- Stürup, G. K. (1953). Sexual offenders and their treatment in Denmark and other Scandinavian countries. *International Review of Criminal Policy* 4, 1–19.
- Stürup, G. K. (1968). Treatment of sexual offenders in Herstedvester Denmark: The rapists. *Acta Psychiatrica Scandinavica Supplement* 204, 1–62.
- Taylor, R. (2000). *A seven-year reconviction study of HMP Grendon Therapeutic Community* (Research Findings No. 115). London: Home Office Research and Statistics Directorate.
- Wille, R. & Beier, K. M. (1989). Castration in Germany. *Annals of Sex Research* 2, 103–133.
- Worling, J. R. & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect* 24, 965–982.
- Ziethen, F. (2002). *Rückfallpräventive Effizienz der sozialtherapeutischen Behandlung von Sexualstraftätern. Evaluation der Sozialtherapie in der JVA Berlin-Tegel* [Relapse preventive efficiency of socialtherapeutic treatment for sexual offenders]. Unpublished master thesis, Freie Universität Berlin.

References

- Alexander, M. A. (1999). Sexual offender treatment efficacy revisited. *Sexual Abuse: A Journal of Research and Treatment* 11, 101–116.
- Aos, S., Phipps, P., Barnoski, R. & Lieb, R. (2001). *The comparative costs and benefits of programs to reduce crime. Version 4.0* (Report No. 01-05-1201). Washington: Washington State Institute for Public Policy.
- Barbaree, H. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity

- of recidivism studies to treatment effects. *Sexual Abuse: Journal of Research and Treatment* 9, 111–128.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C. & Phillips, D. C., et al. (1980). *Toward a reform of program evaluation*. San Francisco, CA: Jossey Bass.
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Science* 587, 49–68.
- Farrington, D. P. & Petrosino, A. (2001). The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science* 578, 35–49.
- Farrington, D. P. & Welsh, B. C. (2003). Family-based prevention of offending: A metaanalysis. *Australian and New Zealand Journal of Criminology* 36, 127–151.
- Fedoroff, J. P. & Moran, B. (1997). Myths and misconceptions about sex offenders. *Canadian Journal of Human Sexuality* 6, 263–276.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In L. V. Hedges (Ed.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.
- Friendship, C., Mann, R. E. & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence* 18, 744–759.
- Furby, L., Weinrott, M. R. & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin* 105, 3–30.
- Gallagher, C. A., Wilson, D. B. & MacKenzie, D. L. (2000). *A meta-analysis of the effectiveness of sex offender treatment programs*. Retrieved 5/23/2001, from <http://www.wam.umd.edu/~wilsondb/papers/sexoffender.pdf>.
- Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology* 63, 802–809.
- Hall, G. C. N. (1996). *Theory-based assessment, treatment, and prevention of sexual aggression*. New York: Oxford University Press.
- Hanson, R. K. & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology* 66, 348–362.
- Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W. D. & Quinsey, V. L., et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment* 14, 169–194.
- Hanson, R. K., Broom, I. & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Sciences* 36, 87–96.
- Hasselblad, V. & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin* 117, 167–178.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Hucker, S. J. & Bain, J. (1990). Androgenic hormones and sexual assault. In W. L. Marshall, D. R. Laws & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 93–102). New York: Plenum Press.
- Langevin, R., Paitich, D., Hucker, S. J., Newman, S., Ramsay, G., Pope, S., Geller, G. & Anderson, C. (1979). The effect of assertiveness training, Provera, and sex of therapist in the treatment of genital exhibitionism. *Journal of Behavior Therapy and Experimental Psychiatry* 10, 275–282.

- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science* 587, 69–81.
- Lipsey, M. W. & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In D. P. Farrington (Ed.), *Serious & violent juvenile offenders: Risk factors and successful interventions* (pp. 313–345). Thousand Oaks, CA: Sage.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
- Lösel, F. (1995). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.), *What works: Reducing reoffending. Guidelines from Practice and Research* (pp. 79–111). Chichester, UK: Wiley.
- Lösel, F. (2000). The efficacy of sexual offender treatment: A brief review of German and international evaluations. In P. J. van Koppen & N. Roos (Eds.), *Rationality, information and progress in law and psychology. In honour of Hans Crombag* (pp. 145–170). Maastricht, NL: Metajuridica Publications.
- Lösel, F. (2001a). Rehabilitation of the offender. In P. B. Baltes (Ed.), *International encyclopedia of the social & behavioral sciences* (Vol. 19, pp. 12988–12993). Amsterdam: Elsevier.
- Lösel, F. (2001b). Evaluating the effectiveness of correctional programs: Bridging the gap between research and practice. In G. A. Bernfeld, D. P. Farrington & A. W. Leschied (Eds.), *Offender rehabilitation in practice. Implementing and evaluating effective programs* (pp. 67–92). Chichester, UK: Wiley.
- Lösel, F. & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Science* 587, 84–109.
- Lösel, F. & Köferl, P. (1989). Evaluation research on correctional treatment in West Germany: A meta-analysis. In H. Wegener, F. Lösel & J. Haisch (Eds.), *Criminal behavior and the justice system* (pp. 334–355). New York: Springer.
- Lösel, F. & Schmucker, M. (2003). *The efficacy of sex offender treatment: A brief synthesis of meta-analyses*. Report for the Campbell Crime and Justice Group (Updated). University of Erlangen–Nuremberg: Institute of Psychology.
- Lösel, F. & Wittmann, W. W. (1989). The relationship of treatment integrity and intensity to outcome criteria. *New Directions for Program Evaluation* 42, 97–108.
- Maletzky, B. M. (1991). *Treating the sexual offender*. Newbury Park, CA: Sage.
- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C. & von Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment* 17, 79–107.
- Marshall, W. L. & McGuire, J. (2003). Effect sizes in the treatment of sexual offenders. *International Journal of Offender Therapy and Comparative Criminology* 47, 653–663.
- McGuire, J. (2002). Integrating findings from research reviews. In J. McGuire (Ed.), *Offender rehabilitation and treatment: Effective programmes and policies to reduce re-offending* (pp. 3–38). Chichester, UK: Wiley.
- McMurrin, M. (Ed.). (2002). *Motivating offenders to change: A guide to enhancing engagement in therapy*. Chichester, UK: Wiley.
- Meyer, W. J. & Cole, C. M. (1997). Physical and chemical castration of sex offenders: A review. *Journal of Offender Rehabilitation* 25, 1–8.
- Meyer, W. J., Cole, C. & Emory, E. (1992). Depo provera treatment for sex offending behavior: An evaluation of outcome. *Bulletin of the American Academy of Psychiatry and the Law* 20, 249–259.

- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods* 3, 354–379.
- Polizzi, D. M., MacKenzie, D. L. & Hickman, L. J. (1999). What works in adult sex offender treatment? A review of prison- and non-prison-based treatment programs. *International Journal of Offender Therapy and Comparative Criminology* 43, 357–374.
- Prentky, R. A., Lee, A. F. S., Knight, R. A. & Cerce, D. (1997). Recidivism rates among child molesters and rapists: A methodological analysis. *Law and Human Behavior* 21, 635–659.
- Quinsey, V. L., Harris, G. T., Rice, M. E. & Lalumière, M. L. (1993). Assessing treatment efficacy in outcome studies of sex offenders. *Journal of Interpersonal Violence* 8, 512–523.
- Rice, M. E. & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. In R. A. Prentky, E. S. Janus & M. C. Seto (Eds.), *Sexually coercive behavior: Understanding and management* (pp. 428–440). New York: New York Academy of Sciences.
- Rösler, A. & Witztum, E. (2000). Pharmacotherapy of paraphilias in the next millennium. *Behavioral Sciences and the Law* 18, 43–56.
- Ruddijs, F. & Timmerman, H. (2000). The Stichting ambulante preventie projecten method: A comparative study of recidivism in first offenders in a Dutch outpatient setting. *International Journal of Offender Therapy and Comparative Criminology* 44, 725–739.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P. & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising. Report to the U.S. Congress*, Washington, DC.
- Weisburd, D., Petrosino, A. & Mason, G. (1993). Design sensitivity in criminal justice experiments. In M. Tonry (Ed.), *Crime and justice* (Vol. 17, pp. 337–379). Chicago: University of Chicago Press.
- Weisburd, D., Lum, C. M. & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science* 578, 50–70.
- Weisburd, D., Lum, C. M. & Yang, S.-M. (2003). When can we conclude that treatments or programs “don't work”? *Annals of the American Academy of Political and Social Science* 587, 31–48.
- White, P., Bradley, C., Ferriter, M. & Hatzipetrou, L. (1998). Managements for people with disorders of sexual preference and for convicted sexual offenders. *The Cochrane Database of Systematic Reviews*. (1998, Issue 4).
- Wille, R. & Beier, K. M. (1989). Castration in Germany. *Annals of Sex Research* 2, 103–133.
- Wilson, D. B. (2001). Meta-analytic methods for criminology. *Annals of the American Academy of Political and Social Science* 578, 71–89.
- Worling, J. R. & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect* 24, 965–982.

About the authors

Friedrich Lösel is professor of psychology and director of the Institute of Psychology at the University of Erlangen-Nuremberg, Germany. His research addresses the origins, development, prevention, and treatment of crime. He is a past president of the European Association of Psychology and Law, president of the German Criminology Society, recipient of the Sellin-Glueck Award of the American Society of Criminology, and fellow of the Academy of Experimental Criminology.

Martin Schmucker is a lecturer and research fellow at the Institute of Psychology at the University of Erlangen, Germany. His areas of research include personality disorders and crime, neuropsychological correlates of antisocial behavior, and the evaluation of treatment and prevention programs.